

Minimum Risk Point Estimation of Gini Index

SHYAMAL KRISHNA DE

School of Mathematical Sciences, National Institute of Science Education and Research, Bhubaneswar,
Odisha, India

(sde@niser.ac.in)

BHARGAB CHATTOPADHYAY *

Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, Texas, USA

(bhargab@utdallas.edu)

Abstract: This paper develops a theory and methodology for estimation of Gini index such that both cost of sampling and estimation error are minimum. Methods in which sample size is fixed in advance, cannot minimize estimation error and sampling cost at the same time. In this article, a purely sequential procedure is proposed which provides an estimate of the sample size required to achieve a sufficiently smaller estimation error and lower sampling cost. Characteristics of the purely sequential procedure are examined and asymptotic optimality properties are proved without assuming any specific distribution of the data. Performance of our method is examined through extensive simulation study.

Keywords: Asymptotic Efficiency; Ratio Regret; Reverse Submartingale; Sequential Point Estimation; Simple Random Sampling.

JEL Classification Code: C400, C440

*Corresponding author: Department of Mathematical Sciences, FO 2.402A, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080, USA, E-mail: bhargab@utdallas.edu.

1. INTRODUCTION

Economic inequality exists in all societies or regions because of the existence of gap in income and wealth among individuals. In order to reduce the gap between the income levels of individuals, government of each and every country devise several economic policies. Periodic evaluation of the effect of economic policies in reducing the income gap between rich and poor is important. There are several inequality indexes in the economic literature. Allison (?) mentioned that among those indices, Gini inequality index is the most widely used measure because it satisfies four basic desirable criteria viz. (i) anonymity, (ii) scale independence, (iii) population independence, and (iv) Pigou-Dalton transfer principle and also Gini index has an easy interpretation and a relation to Lorenz curve.

The most celebrated Gini index, as given in Xu (?), is

$$G_F(X) = \frac{\Delta}{2\mu}, \text{ where } \Delta = E |X_1 - X_2|, \mu = E(X) \quad (1.1)$$

and X_1 & X_2 are two i.i.d. copies of non-negative random variable X . If there are n randomly selected individuals with incomes given by X_1, X_2, \dots, X_n , then an estimator of (1.1) is given by

$$G_n = \frac{\hat{\Delta}_n}{2\bar{X}_n}, \quad (1.2)$$

where \bar{X}_n is the sample mean and $\hat{\Delta}_n$ is the Gini's mean difference (GMD) defined as

$$\hat{\Delta}_n = \binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} |X_{i_1} - X_{i_2}|. \quad (1.3)$$

For continuous evaluation of economic policies implemented by the government, periodic computation of Gini index for the whole country or a region is very important. One source from which Gini index of a region or a country can be calculated is using census data which is typically collected every 10 years. But for estimating the Gini index in intermediate years, data from

annual household survey conducted by government agencies can be used. For instance, National Sample Survey (NSS) in India, European Statistics on Income and Living Conditions in European Union and other agencies conduct household surveys annually or biennially in respective regions or countries. However, many countries, for example Burundi, Chad, Mozambique (as per world bank website), can not afford or do not collect data from households on a relatively large scale atleast biennially.

If household survey data is not available, one has to draw a relatively small sample to estimate the Gini index for that region using appropriate sampling technique. The sampling technique should be chosen depending on the size and socio-economic diversity of the country. For a brief review of several sampling techniques, we refer to Cochran (?). In order to compute Gini index for regions or smaller countries, with lesser social diversity, simple random sampling technique can be used to collect income or expenditure data. There exists literature on statistical inference for inequality indices which is computed from household income or expenditure by means of simple random sampling from the population of interest (e.g., Gastwirth ?, Beach and Davidson, ?, Davidson and Duclos, ?, Xu, ? and Davidson, ?). In this paper, we will use simple random sampling technique to collect income or expenditure data in order to estimate Gini index accurately.

It is well known that error in estimation decreases or in other words accuracy increases, when the sample size increases. This in turn increases the overall cost of sampling. To minimize the cost of sampling, one has to reduce the sample size which in turn may lead to higher estimation error. Thus, a method of estimation should be developed such that both the cost of sampling and the error in estimation are kept as low as possible. In other words, a procedure is required which can act as a trade-off between the estimation error and the sampling cost. To achieve this trade-off, fixed-sample methodologies can not be used, i.e., the sample size should not be fixed in advance. This problem falls in the domain of sequential analysis where it is known as minimum risk point estimation problem. For more details on the literature of sequential analysis, we refer to Ghosh and Sen (?), Ghosh et al. (?), Mukhopadhyay and de Silva (?), and others.

Unlike fixed-sample procedures, sequential procedures do not require sample size to be fixed

in advance. Instead, in a sequential procedure, statistical analysis is continued as the observations are collected. Sampling is terminated according to a pre-defined criterion, also known as stopping rule. Sequential sampling allows the estimation process to finish early requiring small sample size. We are certainly not the first one to suggest sequential methods in econometrics. In fact, there are several articles published in several journals in economics and econometrics which pursued the idea of using sequential or multi-stage inference procedures. Examples include ?, ?, ?, ?, etc.

Below, we provide a brief literature review of some relevant concepts and also our contribution to the literature of statistical inference and economics.

1.1. Literature Review and Our Contributions

The estimator of Gini index in (1.2) involves sample mean and Gini's mean difference which belong to a class of unbiased estimators known as U-statistics. Below, we briefly discuss the literature on U-statistics.

1.1.1. Literature on U-statistics

The theory and practice of U-statistics began with the pioneering papers of Hoeffding (?, ?). In the above papers, Hoeffding derived a general method for obtaining unbiased estimators for a parameter θ associated with an unknown distribution function F . Suppose that X_1, \dots, X_n are *independent and identically distributed* (i.i.d.) random variables from a population with a common distribution function F with an associated parameter $\theta \equiv \theta(F)$, $\theta \in \Theta \subseteq \mathcal{R}$. Then the U-statistic associated with θ is written as follows

$$U \equiv U_n^{(m)} = \binom{n}{m}^{-1} \sum_{(n,m)} g^{(m)}(X_{i_1}, \dots, X_{i_m}),$$

where $\sum_{(n,m)}$ denotes the summation over all possible combinations of indices (i_1, \dots, i_m) such that $1 \leq i_1 < i_2 < \dots < i_m \leq n$, and $m < n$. Here, $g^{(m)}(\cdot)$ is a symmetric kernel of degree m such that $E_F [g^{(m)}(X_1, \dots, X_m)] = \theta(F)$ for all F . Thus both GMD and the sample mean are

U-Statistics with kernels of degree 2 and 1 respectively. Detailed literature on U-statistics can be found in standard textbooks such as Hollander and Wolfe (?), Lee (?), and others.

Apart from being unbiased estimators, U-statistics are reverse martingales with respect to some non-increasing filtration as proven in Lee (p. 119, ?). We exploit the reverse martingale property of U-statistics to derive the asymptotic results in section 3. For more literature on reverse martingales, we refer to classical textbooks on probability theory and stochastic processes such as Loeve (?), Doob (?), and others.

As discussed before, we estimate the Gini index by a sequential method known as minimum risk point estimation (MRPE). This estimation technique is not new in the literature of sequential analysis. Below, we briefly discuss the developments on minimum risk point estimation.

1.1.2. Literature on MRPE

Minimum risk point estimation was first introduced by Robbins (?). He suggested a purely sequential procedure for estimating mean of a normal distribution. Ghosh and Mukhopadhyay (?) generalized this idea to a distribution free scenario and developed a purely sequential procedure for minimum risk point estimation of a population mean. Later, Sen and Ghosh (?) extended the sequential procedure of Ghosh and Mukhopadhyay (?) to accommodate the minimum risk point estimation of any estimable parameter using U-statistics. For more details on MRPE, we refer our readers to Sen (?), Ghosh et al. (?), Mukhopadhyay and de Silva (?), and others.

In minimum risk point estimation problems, a cost function is defined which depends on sample size and error in estimation. In this paper, we will use mean square error (MSE) of Gini index as an error in estimation. We are interested in finding an estimate of unknown optimal sample size which minimizes the asymptotic cost function to estimate Gini index of the population.

1.1.3. Contributions of this paper

Several fixed-sample methods are developed for estimation of Gini index assuming that the incomes from the sampled individuals are independent and identically distributed (i.i.d.). Examples

of such methods can be found in Gastwirth ?, Beach and Davidson, ?, Davidson and Duclos, ?, Xu, ? and Davidson, ?. However, these methods cannot be used for minimum risk point estimation of an inequality index. For a brief overview we refer to ?. In this article, we propose a sequential procedure that yields an asymptotic minimum risk point estimator of Gini index by minimizing the asymptotic risk function defined as a cost function plus a risk term for estimation error. Under very mild assumptions, we prove that the estimated final sample size for our procedure approaches the theoretically optimal sample size that minimizes the cost function. Moreover, we prove that the expected cost for estimating the Gini index using the estimated final sample size is asymptotically close to theoretically expected cost for estimating the Gini index, that is with theoretically optimal sample size. All theoretical results are validated by extensive simulation study.

The remainder of this paper is organized as follows. Section 2 develops a purely sequential procedure which minimizes both the estimation error and the overall sampling cost. Section 3 presents the theoretical properties enjoyed by the proposed sequential procedure. Performance of our method is assessed via simulation study in Section 4. The next section explores the possibility of satisfying stronger asymptotic optimality properties. In Section 6, we provide some concluding remarks. The appendix contains some auxiliary lemmas and detailed proofs of all theoretical results.

2. SEQUENTIAL METHOD OF ESTIMATION

Suppose incomes from n randomly selected individuals are collected. Let the incomes of n persons be X_1, \dots, X_n with a common but unknown distribution function F . The estimator G_n is a biased estimator of the population Gini index, G_F , and $E(G_n - G_F)^2$ is the mean square error (MSE) of G_n . The asymptotic expression for MSE of G_n is given by,

Lemma 2.1. $E(G_F - G_n)^2 = \frac{\xi^2}{n} + O\left(\frac{1}{n^{3/2}}\right)$, where

$$\xi^2 = \frac{\sigma_1^2}{\mu^2} + \frac{\Delta^2 \sigma^2}{4\mu^4} - \frac{\Delta}{\mu^3}(\tau - \mu\Delta), \quad (2.1)$$

$$\sigma_1^2 = V [E (|X_1 - X_2| \mid X_1)] , \tau = E (X_1 | X_1 - X_2|) , \text{ and } \sigma^2 = V(X),$$

provided $E(X_1^{12})$ and $E(X_1^{-20})$ exist.

The proof of the lemma is given in the appendix. If the sample size is large, we receive more and more information about G_F and, therefore, expect the squared error loss $(G_F - G_n)^2$ due to estimation to be small. However, higher sample size leads to higher sampling cost. Therefore, it is desirable to consider a loss function that takes into account both loss due to error in estimation and the sampling cost. Suppose c is the known cost of sampling each observation. Our goal is to find an estimation procedure which minimizes both the MSE and also the sampling cost. We define a cost function depending on the MSE and the cost of sampling, also known as the risk function, as

$$R_n(G_F) = AE(G_F - G_n)^2 + cn. \quad (2.2)$$

Here, A is a known positive constant and is expressed in monetary terms which represents the weight assigned by the researchers or analysts regarding the probable cost per unit squared error loss due to estimation. Thus, the first term $AE(G_F - G_n)^2$ represents the loss in estimating G_F by G_n , and the second term cn represents the cost of sampling n observations. The risk function thus gives the expected cost of estimating G_F using the estimator G_n based on incomes from n individuals. Using the asymptotic expression of MSE of G_n expressed in (2.1), the fixed-sample size risk defined in (2.2) becomes

$$R_n(G_F) = A \frac{\xi^2}{n} + cn + O\left(\frac{1}{n^{3/2}}\right). \quad (2.3)$$

Thus, (2.3) gives the expected cost or the risk, to estimate the unknown value of the population Gini index using G_n based on n observations. Our goal is to find the sample size for which the approximate expected cost (ignoring the $O\left(\frac{1}{n^{3/2}}\right)$ term) defined in (2.3), i.e., $h(n) = A \frac{\xi^2}{n} + cn$ is minimized for all distributions that satisfy the conditions of lemma 2.1.

Considering n as a non-negative continuous variable, the strictly convex function $h(n)$ can

be minimized at $n = n_c \left(= \sqrt{\frac{A}{c}} \xi \right)$. Thus n_c is the required optimal sample size that should be collected using simple random sampling from the population in order to minimize the expected cost to estimate G_F . Thus the approximate expected cost of estimating the Gini index using a sample of size n_c or the asymptotic minimum risk is

$$R_{n_c}^*(G_F) = A \frac{\xi^2}{n_c} + cn_c = 2cn_c. \quad (2.4)$$

If the parameter ξ were known in advance, one could simply collect a sample of size n_c which is the minimum sample size to attain the asymptotic minimum risk. Since ξ is not known, we need to collect samples in at least two stages where the first stage is to estimate ξ and n_c based on a pilot sample. In fact, Dantzig (?) proved that fixed-sample procedures cannot minimize the risk in (2.3), not even asymptotically. Therefore, we propose a purely sequential procedure that yields minimum risk at least asymptotically.

Since ξ is unknown, we first provide an estimator of ξ that is strongly consistent. The estimator of ξ is based on U-statistics and can also be found in Xu (?), and Sproule (?). Proceeding along the lines of Sproule (?), let us define a U-statistic, for each $j = 1, 2, \dots, n$,

$$\widehat{\Delta}_n^{(j)} = \binom{n-1}{2}^{-1} \sum_{T_j} |X_{i_1} - X_{i_2}|,$$

where $T_j = \{(i_1, i_2) : 1 \leq i_1 < i_2 \leq n \text{ and } i_1, i_2 \neq j\}$. Also, define $W_{jn} = n\widehat{\Delta}_n - (n-2)\widehat{\Delta}_n^{(j)}$, for $j = 1, \dots, n$, and $\overline{W}_n = n^{-1} \sum_{j=1}^n W_{jn}$. According to Sproule (?), a strongly consistent estimator of $4\sigma_1^2$ is

$$s_{wn}^2 = (n-1)^{-1} \sum_{i=1}^n (W_{jn} - \overline{W}_n)^2.$$

Using Xu (2007),

$$\widehat{\tau}_n = \frac{2}{n(n-1)} \sum_{(n,2)} \frac{1}{2} (X_{i_1} + X_{i_2}) |X_{i_1} - X_{i_2}|$$

is an estimator of τ . Let S_n^2 be the sample variance. Thus, the estimator of ξ^2 is

$$V_n^2 = \frac{\hat{\Delta}_n^2 S_n^2}{4\bar{X}_n^4} - \frac{\hat{\Delta}_n}{\bar{X}_n^3} \hat{\tau}_n + \frac{\hat{\Delta}_n^2}{\bar{X}_n^2} + \frac{s_{wn}^2}{4\bar{X}_n^2}. \quad (2.5)$$

Using Sproule (1969) and theorem 3.2.1 of Sen (1981, p. 50), we conclude that V_n^2 is a strongly consistent estimator of ξ^2 .

We outline the purely sequential estimation procedure of the Gini Index of the population as follows:

Step 1: In the first step, often called the pilot sample step, incomes from a sample of m individuals are collected. This sample is called the pilot sample. Based on this pilot sample of size m , an estimate of ξ^2 obtained by computing V_m^2 . Check the condition, $m \geq \sqrt{\frac{A}{c}} V_m$. If $m < \sqrt{\frac{A}{c}} V_m$ then go to the next step. Otherwise, if $m \geq \sqrt{\frac{A}{c}} V_m$, then stop sampling and set the the final sample size equal to m .

Step 2: Obtain income from one randomly selected individuals. Update the estimate of ξ^2 and verify the condition based on $m + 1$ observations. If $m + 1 \geq \sqrt{\frac{A}{c}} V_{m+1}$ stop further sampling and set the final sample size equal to $m + 1$. If $m + 1 < \sqrt{\frac{A}{c}} (V_{m+1})$ then continue the sampling process by sampling 1 more individuals and simultaneously update the condition.

The sampling process is continued until the updated condition is satisfied.

Formally, we define the stopping rule N , for every $c > 0$, as

$$N \equiv N(c) \text{ is the smallest integer } n(\geq m) \text{ such that } n \geq \sqrt{\frac{A}{c}} V_n. \quad (2.6)$$

Here, m is the initial or pilot sample size. In some extreme situations, the estimator V_n may be very small which may cause our procedure to stop too early. To avoid this problem, we propose a slightly modified stopping rule N_c as

$$N_c \text{ is the smallest integer } n(\geq m) \ni n \geq \sqrt{\frac{A}{c}} (V_n + n^{-\gamma}), \quad (2.7)$$

where $\gamma \in (0, 0.5)$ is a suitable constant. The inclusion of the term $n^{-\gamma}$ ensures that we do not stop too early due to small value of V_n .

3. THEORETICAL RESULTS

For a given cost c per observation, the risk or the expected cost for estimating the Gini index G_F using an estimator based on the final sample size N_c is given by

$$R_{N_c}(G_F) = AE(G_F - G_{N_c})^2 + cE(N_c). \quad (3.1)$$

Thus, the estimator G_{N_c} is asymptotically minimum risk point estimator (AMRPE) if the ratio regret is asymptotically 1, i.e., if

$$\lim_{c \rightarrow 0} R_{N_c}(G_F) / R_{n_c}(G_F) = 1. \quad (3.2)$$

In other words, estimator G_{N_c} is AMRPE (refer Sen, 1981) if the expected cost for estimating the Gini index G_F using an estimator based on the final sample size N_c is asymptotically close to expected cost for estimating G_F using the optimal sample size, n_c . In decision theoretic framework, the ratio in (3.2) is known as ratio regret which is the ratio between the actual payoff and the minimum payoff due to some optimal strategy (Loomes and Sugden, ?).

Before discussing the asymptotic optimality properties of our method, we prove in the following lemma that if observations are collected using (2.7), sampling will stop at some finite time with probability one.

Lemma 3.1. *Under the assumption that $\xi < \infty$, for any $c > 0$, the stopping time N_c is finite, i.e., $P(N_c < \infty) = 1$.*

Proof of this lemma is given in Appendix. This lemma is very crucial for any sequential procedure because it assures that the practitioner will not need to sample indefinitely. Below we provide the main theorem related to the asymptotic optimality properties of our procedure.

Theorem 3.1. *The stopping rule (2.7) yields:*

(i) $N_c/n_c \rightarrow 1$ almost surely as $c \downarrow 0$.

(ii) $E(N_c/n_c) \rightarrow 1$ as $c \downarrow 0$. [Asymptotic First-order Efficiency]

(iii) If $\gamma \in (0, \frac{1}{2})$, $R_{N_c}(G_F)/R_{n_c}(G_F) \rightarrow 1$ as $c \downarrow 0$. [Asymptotic First-order Risk Efficiency]

provided, $E(X^{16})$ and $E(X^{-24})$ exist.

Proof. Proof of this theorem is technical and, therefore, it is given in Appendix. □

The parts (i) and (ii) of this theorem imply that the final sample size of our procedure is asymptotically same as the minimum sample size required to minimize the asymptotic risk defined in (2.3). The part (iii) proves that the risk attained by our procedure is asymptotically same as the minimum risk. Therefore, the Gini index estimator G_{N_c} is indeed AMRPE. The optimality properties in part (ii) and (iii) are well known in the sequential literature as asymptotic first-order efficiency and asymptotic first-order risk efficiency respectively (see Mukhopadhyay and de Silva, 2009). Theorem 3.1 also holds for the stopping rule defined in (2.6).

4. PERFORMANCE VIA SIMULATIONS

In this section, we evaluate performance of our estimation strategy for moderate sample size (i.e., c is small but not too small) via simulation study.

To implement the sequential procedure in (2.6), we fix $c = 0.1$, $A = 50000$, and the pilot sample size $m = 10$. The results in Table 1 and 2 are based on random samples from three income distributions: exponential (rate = 5), gamma (shape = 2.649, rate = 0.84), and log-normal (mean = 2.185, sd = 0.562). Number of replications used in all Monte carlo simulations is 5000. Table 1 compares the true values of the parameters σ_1^2 , τ , and ξ^2 with their estimated values based on the final sample size N . $s(s_{wN}^2)$, $s(\hat{\tau}_N)$, and $s(V_N^2)$ represent the standard errors of the estimators s_{wN}^2 , $\hat{\tau}_N$, and V_N^2 respectively.

Table 1. Estimated sample variances and covariances

Distribution	$\overline{s_{wN}^2}$ $\mathbf{s}(s_{wN}^2)$	$4\sigma_1^2$	$\overline{\hat{\tau}_N}$ $\mathbf{s}(\hat{\tau}_N)$	τ	$\overline{V_N^2}$ $\mathbf{s}(V_N^2)$	ξ^2
Exponential	0.0521 0.0002	0.0532	0.0596 0.0001	0.06000	0.0843 0.0002	0.0833
Gamma	3.4172 0.0157	3.5036	7.8110 0.0147	7.8205	0.0463 0.0001	0.0468
Lognormal	52.11274 0.1173	52.8108	84.9292 0.0694	85.2236	0.0498 0.00009	0.0526

Table 2. Estimated average final sample size and the ratio regret

Distribution	\overline{N} $\mathbf{s}(\overline{N})$	n_c	\overline{N}/n_c	$\max(N)$	$\overline{\tau}_N$ $\mathbf{s}(\overline{\tau}_N)$	$\frac{\overline{\tau}_N}{R^*_{n_c}}$
Exponential	205.4111 0.2378	204.08	1.0065	319	40.9317 0.0474	1.0028
Gamma	152.19 0.1970	152.97	0.9949	239	30.2765 0.0391	0.9904
Lognormal	162.3504 0.1483	163.10	0.9954	228	152.07 0.1958	0.9919

Table 1 shows that the average values of the estimators are close to the true values of the parameters and, therefore, it indicates that $s_{wN}^2 \rightarrow 4\sigma_1^2$, $\hat{\tau}_N \rightarrow \tau$ and $V_N^2 \rightarrow \xi^2$ as $c \downarrow 0$.

Table 2 presents the average final sample size \bar{N} (estimates $E(N)$), the maximum sample size $\max(N)$ from 5000 replications, and the average risk \bar{r}_N (estimates $R_N(G_F)$) obtained from the sample of size N . Moreover, $s(\bar{N})$ and $s(\bar{r}_N)$ represent the standard errors of \bar{N} and \bar{r}_N respectively. Table 2 shows that the average sample size \bar{N} is almost the same as the optimal sample size n_c . Therefore, on average, our procedure requires only the minimum sample size n_c . The last column of Table 2 illustrates that, on average, the cost for estimating the Gini index G_F using an estimator based on the estimated final sample size is asymptotically close to expected cost for estimating G_F using the optimal sample size, n_c , or in other words, the ratio regret is very close to 1. This implies that the risk incurred by our method is almost the same as the minimum possible risk $R_{n_c}^*$ defined in (2.4). Thus, we find that the proposed sequential procedure performs remarkably well for the above mentioned income distributions.

5. EXTENSIONS AND DISCUSSIONS

5.1. Exploring Asymptotic Second-Order Efficiency

In sequential point estimation literature, a stopping rule N_c is known as *asymptotically second-order efficient* (see Ghosh and Mukhopadhyay, ?) if the difference between the expected final sample size $E(N_c)$ and the theoretically optimum fixed-sample size n_c is asymptotically bounded, i.e., if $E(N_c) - n_c$ is bounded as $c \downarrow 0$. Clearly, if a sequential method is second-order efficient, it is first-order efficient as well. However, the converse is not necessarily true. We explore this second-order efficiency property via Monte Carlo simulations. Under the same scenario as in Tables 1 and 2, we apply our method and estimate the difference $E(N_c) - n_c$ based on 500 replications. We repeat this process 10 times and present 10 observed values of $\bar{N} - n_c$ each estimating $E(N_c) - n_c$. Table 3 shows that the differences $E(N_c) - n_c$ are quite small for all three distributions. Therefore, simulation study strongly indicates that the proposed sequential procedure is asymptotically

Table 3. Estimated Values of $E[N_c] - n_c$

Distribution	$E[N_c] - n_c$				
Exponential	1.9800	0.7660	1.5520	2.600	1.0380
	1.1560	0.6760	1.5100	0.7020	1.9020
Gamma	-1.178	-0.98	-0.274	-1.338	-1.2
	-0.834	-0.438	-0.686	-0.282	-0.626
Lognormal	-0.7211	-0.9591	-0.8331	-0.7951	0.6771
	-0.6611	-0.1091	-1.2671	-1.2211	-0.2031

second-order efficient.

5.2. Exploring Asymptotic Second-Order Risk Efficiency

In sequential point estimation literature, a stopping rule N_c is known as *asymptotically second-order risk efficient* (see Ghosh and Mukhopadhyay, ?) if the difference regret, i.e., $R_{N_c}(G_F) - R_{n_c}(G_F)$ is asymptotically bounded. This property implies asymptotic first-order risk efficiency. We explore this second-order risk efficiency property via Monte Carlo simulations. For each of the three distributions in Table 4, 10 observed values of $\bar{r}_{N_c} - R_{n_c}$ are presented, each estimating $R_{N_c}(G_F) - R_{n_c}(G_F)$. Table 4 shows that the differences $R_{N_c}(G_F) - R_{n_c}(G_F)$ are quite small for all three distributions. Monte Carlo simulations strongly indicates that the proposed sequential procedure is asymptotically second-order risk efficient.

6. CONCLUDING REMARKS

The Gini index or Gini concentration is a very popular measure of inequality. It is well known that error in estimation of Gini index decreases when the sample size increases. This inflates the overall cost of sampling. In order to compute Gini index for a region or a smaller country with lesser diversity at a specific point of time, we develop a procedure which computes the final sample size needed to minimize both the error of estimation as well as the cost of sampling via

Table 4. Estimated Values of $R_{N_c}(G_F) - R_{n_c}(G_F)$

Distribution	$\bar{r}_{N_c} - R_{n_c}$ values				
Exponential	0.2397	-0.0016	0.1568	0.3799	0.0619
	0.0796	-0.0150	0.1564	-0.0165	0.2302
Gamma	-0.3733	-0.3387	-0.1901	-0.4009	-0.3769
	-0.3048	-0.2218	-0.2768	-0.1933	-0.2547
Lognormal	-0.3080	-0.2878	-0.2796	-0.2480	-0.2494
	-0.3704	-0.3587	-0.1540	-0.2625	-0.1379

simple random sampling technique.

Without assuming any specific distribution for the data, we showed that the average final sample size using our procedure approaches the unknown optimal sample size that minimizes the cost function. Moreover, we proved that the expected cost for estimating the Gini index using the estimated final sample size is asymptotically close to the expected cost for estimating the Gini index using the unknown optimal sample size. Thus, based on the results mentioned above, we conclude that the proposed sequential estimation strategy is remarkably efficient in reducing both sampling cost and estimation error.

7. APPENDIX: AUXILIARY RESULTS AND PROOFS

7.1. Proof of Lemma 3.1

Note that V_n is strongly consistent estimator of ξ . Therefore, for any fixed $c > 0$,

$$\begin{aligned}
P(N_c > \infty) &= \lim_{n \rightarrow \infty} P(N_c > n) \\
&= \lim_{n \rightarrow \infty} P\left(n < \sqrt{A/c}(V_n + n^{-\gamma})\right) = 0.
\end{aligned}$$

The last equality is obtained since $V_n \rightarrow \xi$ almost surely as $n \rightarrow \infty$. This completes the proof.

7.2. Lemmas to Prove The Main Result

This section is dedicated to prove some lemmas that are essential to establish the main theorem

3.1. First, we introduce few notations. Note from (2.7) that $N_c \geq \sqrt{\frac{A}{c}} N_c^{-\gamma}$, i.e., $N_c \geq \left(\frac{A}{c}\right)^{\frac{1}{2(1+\gamma)}}$ with probability 1. For fixed $\epsilon, \gamma > 0$, define

$$n_{1c} = \left(\frac{A}{c}\right)^{\frac{1}{2(1+\gamma)}}, \quad n_{2c} = n_c(1 - \epsilon), \quad \text{and} \quad n_{3c} = n_c(1 + \epsilon), \quad \text{where} \quad n_c = \sqrt{\frac{A}{c}} \xi. \quad (7.1)$$

Suppose $\mathbf{X}_{(n)}$ denotes the n dimensional vector of order statistics from the sample X_1, \dots, X_n , and \mathcal{F}_n is the σ -algebra generated by $(\mathbf{X}_{(n)}, X_{n+1}, X_{n+2}, \dots)$. By Lee (1990), $\{\bar{X}_n, \mathcal{F}_n\}$, $\{S_n^2, \mathcal{F}_n\}$, $\{\hat{\tau}_n, \mathcal{F}_n\}$, $\{\hat{\Delta}_n, \mathcal{F}_n\}$, and their convex functions are all reverse submartingales. Using reverse submartingale properties of U-statistics, we prove the following maximal inequality for sample Gini's mean difference.

Lemma 7.1. *If nonnegative i.i.d. random variables X_1, \dots, X_n are from the distribution F such that $E(X_1^{\max(2r, p)}) < \infty$ for some positive integers r and p , then for any $k > 0$,*

$$P\left(\max_{n_{1c} \leq n \leq n_{2c}} \left|\hat{\Delta}_n^2 - \Delta^2\right| \geq k\right) \leq O(n_{1c}^{-r/2}) + O(n_{1c}^{-p/2}) \quad \text{as } c \downarrow 0.$$

Proof. Note that

$$\begin{aligned} \left|\hat{\Delta}_n^2 - \Delta^2\right| &= \left|\left(\hat{\Delta}_n^2 - \Delta^2\right) I(\hat{\Delta}_n > \Delta) + \left(\hat{\Delta}_n^2 - \Delta^2\right) I(\hat{\Delta}_n < \Delta)\right| \\ &\leq \left(\hat{\Delta}_n^2 - \Delta^2\right)^+ + 2\Delta \left|\hat{\Delta}_n - \Delta\right| I(\hat{\Delta}_n < \Delta). \end{aligned} \quad (7.2)$$

Here, the notation x^+ is used to mean $\max(x, 0)$. Therefore,

$$P\left(\max_{n_{1c} \leq n \leq n_{2c}} \left|\hat{\Delta}_n^2 - \Delta^2\right| \geq k\right) \leq P\left(\max_{n_{1c} \leq n \leq n_{2c}} \left(\hat{\Delta}_n^2 - \Delta^2\right)^+ \geq \frac{k}{2}\right) + P\left(\max_{n_{1c} \leq n \leq n_{2c}} \left|\hat{\Delta}_n - \Delta\right| \geq \frac{k}{4\Delta}\right).$$

Since $\left(\hat{\Delta}_n^2 - \Delta^2\right)^+$ and $\left|\hat{\Delta}_n - \Delta\right|$ are reverse submartingales, using maximal inequality for reverse

submartingales (Ghosh et al. 1997), we write

$$\begin{aligned}
P\left(\max_{n_{1c} \leq n \leq n_{2c}} \left| \hat{\Delta}_n^2 - \Delta^2 \right| \geq k\right) &\leq \left(\frac{2}{k}\right)^r E\left[\left(\left(\hat{\Delta}_{n_{1c}}^2 - \Delta^2\right)^+\right)^r\right] + \left(\frac{4\Delta}{k}\right)^p E\left[\left|\hat{\Delta}_{n_{1c}} - \Delta\right|^p\right] \\
&\leq \left(\frac{2}{k}\right)^r \left\{E\left[\left(\hat{\Delta}_{n_{1c}} - \Delta\right)^{2r}\right] E\left[\left(\hat{\Delta}_{n_{1c}} + \Delta\right)^{2r}\right]\right\}^{\frac{1}{2}} + O\left(n_{1c}^{-\frac{p}{2}}\right) \\
&\leq O\left(n_{1c}^{-r/2}\right) + O\left(n_{1c}^{-p/2}\right).
\end{aligned}$$

The last two inequalities are obtained by Cauchy–Schwarz inequality and lemma 2.2 of Sen and Ghosh (1981). The moment conditions of this lemma are needed to ensure that all expectaions exist in the last three inequalities. \square

Lemma 7.2. *Let \bar{X}_n be the sample mean based on nonnegative i.i.d. observations X_1, \dots, X_n . For $r \geq 1$, $E\left(\bar{X}_n^{-r}\right) \leq E\left(X_1^{-r}\right)$.*

Proof. Note that $\bar{X}_n \geq (\prod_{i=1}^n X_i)^{1/n}$ as the observations are nonnegative. Therefore,

$$E\left(\bar{X}_n^{-r}\right) \leq E\left[\left(\prod_{i=1}^n \frac{1}{X_i}\right)^{r/n}\right] = \left\{E\left[\left(\frac{1}{X_1}\right)^{r/n}\right]\right\}^n. \quad (7.3)$$

The last equality is due to the i.i.d. property of the observations. We know that $\{E(|X|^p)\}^{1/p}$ is a nondecreasing function of p for $p > 0$. Applying this result with $p = 1/n \geq 1$ in (7.3), we complete the proof. \square

Lemma 7.3. *Suppose that nonnegative i.i.d. random variables X_1, \dots, X_n are observed from the distribution F such that $E(X_1)^{4p}$ and $E(X_1)^{-\max\{4p, 2p(r-1)\}}$ exist for some positive integers r and p . Then, for any $k > 0$,*

$$P\left(\max_{n_{1c} \leq n \leq n_{2c}} \left|\frac{1}{\bar{X}_n^r} - \frac{1}{\mu^r}\right| \geq k\right) \leq O(n_{1c}^{-p/2}) \text{ as } c \downarrow 0.$$

Proof. By Taylor expansion of $\bar{X}_n^{-r} = \frac{1}{\mu^r} (1 + (\bar{X}_n - \mu)/\mu)^{-r}$, we have

$$\left| \left(\frac{1}{\bar{X}_n^r} - \frac{1}{\mu^r} \right) I \left(\frac{1}{\bar{X}_n} < \frac{1}{\mu} \right) \right| = \frac{1}{\mu^r} \left| \left\{ -\frac{r}{\mu} (\bar{X}_n - \mu) + \frac{r(r+1)}{2\mu^2} \frac{(\bar{X}_n - \mu)^2}{z^{r+2}} \right\} I \left(\frac{1}{\bar{X}_n} < \frac{1}{\mu} \right) \right|,$$

where $z \in [1, \bar{X}_n/\mu]$. Since $z^{-(r+2)} I(\bar{X}_n^{-1} < \mu^{-1}) \leq 1$, proceeding along the lines of (7.2)

$$\begin{aligned} \left| \frac{1}{\bar{X}_n^r} - \frac{1}{\mu^r} \right| &= \left| \left(\frac{1}{\bar{X}_n^r} - \frac{1}{\mu^r} \right) I \left(\frac{1}{\bar{X}_n} \geq \frac{1}{\mu} \right) + \left(\frac{1}{\bar{X}_n^r} - \frac{1}{\mu^r} \right) I \left(\frac{1}{\bar{X}_n} < \frac{1}{\mu} \right) \right| \\ &\leq \left(\frac{1}{\bar{X}_n^r} - \frac{1}{\mu^r} \right)^+ + \frac{r}{\mu^{r+1}} |\bar{X}_n - \mu| + \frac{r(r+1)}{2\mu^{r+2}} (\bar{X}_n - \mu)^2. \end{aligned} \quad (7.4)$$

Let $U_{1n} = \left(\frac{1}{\bar{X}_n^r} - \frac{1}{\mu^r} \right)^+$, $U_{2n} = \frac{r}{\mu^{r+1}} |\bar{X}_n - \mu|$, and $U_{3n} = \frac{r(r+1)}{2\mu^{r+2}} (\bar{X}_n - \mu)^2$. Using (7.4), we can write

$$\begin{aligned} P \left(\max_{n_{1c} \leq n \leq n_{2c}} \left| \frac{1}{\bar{X}_n^r} - \frac{1}{\mu^r} \right| \geq k \right) &\leq P \left(\max_{n_{1c} \leq n \leq n_{2c}} U_{1n} \geq \frac{k}{3} \right) + P \left(\max_{n_{1c} \leq n \leq n_{2c}} U_{2n} \geq \frac{k}{3} \right) \\ &\quad + P \left(\max_{n_{1c} \leq n \leq n_{2c}} U_{3n} \geq \frac{k}{3} \right). \end{aligned} \quad (7.5)$$

Since $\left(\frac{1}{\bar{X}_n^r} - \frac{1}{\mu^r} \right)$ is a reverse submartingale and $f(x) = x^+$ is a non-decreasing convex function of x , U_{1n} is a reverse submartingale. Therefore, using maximal inequality for reverse submartingales

$$\begin{aligned} P \left(\max_{n_{1c} \leq n \leq n_{2c}} U_{1n} \geq \frac{k}{3} \right) &\leq \left(\frac{3}{k} \right)^p E \left[\left(\frac{1}{\bar{X}_{n_{1c}}^r} - \frac{1}{\mu^r} \right)^+ \right]^p \\ &\leq \left(\frac{3}{k} \right)^p E \left[\left(\frac{1}{\bar{X}_{n_{1c}}^r} - \frac{1}{\mu^r} \right) \left(\frac{1}{\bar{X}_{n_{1c}}^{r-1}} + \frac{1}{\mu \bar{X}_{n_{1c}}^{r-2}} + \dots + \frac{1}{\mu^{r-1}} \right) I(\bar{X}_{n_{1c}} < \mu) \right]^p \\ &\leq \left(\frac{3}{k} \right)^p r^p E \left[\left(\frac{1}{\bar{X}_{n_{1c}}^r} - \frac{1}{\mu^r} \right)^p \bar{X}_{n_{1c}}^{-p(r-1)} \right] \\ &\leq \left(\frac{3r}{k} \right)^p \left\{ E[(\bar{X}_{n_{1c}} - \mu)^{4p}] E \left[\left(\frac{1}{\mu \bar{X}_{n_{1c}}} \right)^{4p} \right] \right\}^{\frac{1}{4}} \left\{ E \left(\frac{1}{\bar{X}_{n_{1c}}^{2p(r-1)}} \right) \right\}^{\frac{1}{2}} \\ &\leq O(n_{1c}^{-p/2}). \end{aligned} \quad (7.6)$$

The last two inequalities are obtained by using Cauchy–Schwarz inequality and lemma 2.2 of Sen

and Ghosh (1981). Due to lemma 7.2, existence of $E(X_1)^{-\max\{4p, 2p(r-1)\}}$ ensures the existence of $E\left[\left(\frac{1}{\bar{X}_{n_{1c}}}\right)^{4p}\right]$ and $E\left[\left(\frac{1}{\bar{X}_{n_{1c}}}\right)^{2p(r-1)}\right]$. Since $|\bar{X}_n - \mu|$ and $(\bar{X}_n - \mu)^2$ are reverse submartingales, we can write

$$P\left(\max_{n_{1c} \leq n \leq n_{2c}} U_{2n} \geq \frac{k}{3}\right) \leq \left(\frac{3r}{k\mu^{r+1}}\right)^{2p} E(\bar{X}_{n_{1c}} - \mu)^{2p} \leq O(n_{1c}^{-p}), \quad (7.7)$$

$$P\left(\max_{n_{1c} \leq n \leq n_{2c}} U_{3n} \geq \frac{k}{3}\right) \leq \left(\frac{3r(r+1)}{2k\mu^{r+2}}\right)^p E(\bar{X}_{n_{1c}} - \mu)^{2p} \leq O(n_{1c}^{-p}). \quad (7.8)$$

Apply (7.6), (7.7), and (7.8) in (7.5) to complete the proof. \square

Lemma 7.4. Suppose that nonnegative i.i.d. observations X_1, \dots, X_n are such that $E(X_1^{4r})$ and $E(X_1^{-6r})$ exist for some $r \geq 1$. For any $\epsilon \in (0, 1)$ and $\gamma > 0$,

$$(i) \ P(N_c \leq n_c(1 - \epsilon)) = O\left(n_{1c}^{-\frac{r}{2}}\right) = O\left(c^{\frac{r}{4(1+\gamma)}}\right) \text{ as } c \downarrow 0,$$

$$(ii) \ P(N_c \geq n_c(1 + \epsilon)) = O\left(n_{1c}^{-\frac{r}{2}}\right) = O\left(c^{\frac{r}{4(1+\gamma)}}\right) \text{ as } c \downarrow 0.$$

Proof. Using the definition of stopping rule N_c in (2.7) and (7.1), we have

$$\begin{aligned} P(N_c \leq n_{2c}) &\leq P\left(n > \sqrt{\frac{A}{c}} V_n \text{ for some } n \in [n_{1c}, n_{2c}]\right) \\ &\leq P\left(V_n^2 \leq \left(\frac{c}{A}\right) n_{2c}^2 \text{ for some } n \in [n_{1c}, n_{2c}]\right) \\ &\leq P(|V_n^2 - \xi^2| \geq \xi^2 \epsilon (2 - \epsilon) \text{ for some } n \in [n_{1c}, n_{2c}]) \\ &\leq P\left(\max_{n_{1c} \leq n \leq n_{2c}} \{|V_{1n}| + |V_{2n}| + |V_{3n}| + |V_{4n}|\} \geq \xi^2 \epsilon (2 - \epsilon)\right), \end{aligned} \quad (7.9)$$

where $V_{1n} = \left(\frac{\hat{\Delta}_n^2}{4\bar{X}_n^4} S_n^2 - \frac{\Delta^2}{4\mu^4} \sigma^2\right)$, $V_{2n} = \left(\frac{\hat{\Delta}_n}{\bar{X}_n^3} \hat{\tau}_n - \frac{\Delta}{\mu^3} \tau\right)$, $V_{3n} = \left(\frac{\hat{\Delta}_n^2}{\bar{X}_n^2} - \frac{\Delta^2}{\mu^2}\right)$, and $V_{4n} = \left(\frac{s_{wn}^2}{4\bar{X}_n^2} - \frac{\sigma_1^2}{\mu^2}\right)$.

Let $k = \xi^2 \epsilon (2 - \epsilon)$. Then, (7.9) can be written as $P(N_c \leq n_{2c}) \leq P_1 + P_2 + P_3 + P_4$, where

$$P_i = P\left(\max_{n_{1c} \leq n \leq n_{2c}} |V_{in}| \geq \frac{k}{4}\right), \quad \text{for } i = 1, 2, 3, 4.$$

First, let us find an upper bound of P_1 . Let $T_{1n} = \left(\hat{\Delta}_n^2 - \Delta^2\right)$, $T_{2n} = (S_n^2 - \sigma^2)$, and $T_{3n} =$

$\left(\frac{1}{4\bar{X}_n^4} - \frac{1}{4\mu^4}\right)$. Note that

$$V_{1n} = T_{1n}T_{2n}T_{3n} + \Delta^2 T_{2n}T_{3n} + \sigma^2 T_{1n}T_{3n} + \frac{1}{\mu^4} T_{1n}T_{2n} + \frac{\sigma^2}{\mu^4} T_{1n} + \frac{\Delta^2}{\mu^4} T_{2n} + \Delta^2 \sigma^2 T_{3n}, \quad (7.10)$$

Let us consider the first term in the summation of (7.10) and state the following inequalities.

$$\begin{aligned} P\left(\max_{n_{1c} \leq n \leq n_{2c}} |T_{1n}T_{2n}T_{3n}| \geq \frac{k}{28}\right) &\leq \sum_{i=1}^3 P\left(\max_{n_{1c} \leq n \leq n_{2c}} |T_{in}| \geq \left(\frac{k}{28}\right)^{\frac{1}{3}}\right) \\ &\leq O(n_{1c}^{-r}) + O(n_{1c}^{-r}) + O(n_{1c}^{-r/2}) = O(n_{1c}^{-r/2}). \end{aligned} \quad (7.11)$$

The asymptotic orders in (7.11) are obtained by using lemma 7.1, maximal inequality for reverse martingales (Lee, p. 112, 1990), lemma 2.2 of Sen and Ghosh (1981), and lemma 7.3. The conditions of lemma 7.4 are also used in (7.11). Following the same argument as above, one can show that the asymptotic order of probability of large deviations (as in (7.11)) corresponding to the remaining six terms in the summation of (7.10) are either $O(n_{1c}^{-r})$ or $O(n_{1c}^{-r/2})$. Therefore,

$$P_1 = P\left(\max_{n_{1c} \leq n \leq n_{2c}} |V_{1n}| \geq \frac{k}{4}\right) \leq O(n_{1c}^{-r/2}). \quad (7.12)$$

Note that all the estimators in V_{2n} and V_{3n} are U-statistics as we had in the case of V_{1n} . So, following similar arguments as in the proof of (7.12), one can show that both P_2 and P_3 are $O(n_{1c}^{-r/2})$ as $c \downarrow 0$.

To work with P_4 , we note that the expression of V_{4n} involves s_{wn}^2 which is not a U-statistics. Therefore, arguments given in the case of P_1 - P_3 may not work without additional result. Following the proof of lemma 3.1 of Sen and Ghosh (1981) and noting that $E(X_1^{4r}) < \infty$ for $r \geq 1$,

$$P\left(\max_{n_{1c} \leq n \leq n_{2c}} \left|\frac{s_{wn}^2}{4} - \sigma_1^2\right| \geq K\right) \leq O(n_{1c}^{-r}), \quad \text{for any positive constant } K. \quad (7.13)$$

Noting that $V_{4n} = W_{1n}W_{2n} + \sigma_1^2 W_{2n} + \mu^{-2} W_{1n}$, where $W_{1n} = \left(\frac{s_{wn}^2}{4} - \sigma_1^2\right)$ and $W_{2n} = \left(\frac{1}{\bar{X}_n^2} - \frac{1}{\mu^2}\right)$,

$$\begin{aligned} P_4 &\leq P\left(\max_{n_{1c} \leq n \leq n_{2c}} |W_{1n}W_{2n}| \geq \frac{k}{12}\right) + P\left(\max_{n_{1c} \leq n \leq n_{2c}} |W_{2n}| \geq \frac{k}{12\sigma_1^2}\right) + P\left(\max_{n_{1c} \leq n \leq n_{2c}} |W_{1n}| \geq \frac{k\mu^2}{12}\right) \\ &\leq \sum_{i=1}^2 P\left(\max_{n_{1c} \leq n \leq n_{2c}} |W_{in}| \geq \sqrt{\frac{k}{12}}\right) + O(n_{1c}^{-r/2}) + O(n_{1c}^{-r}) \leq O(n_{1c}^{-r/2}). \end{aligned} \quad (7.14)$$

The asymptotic orders in (7.14) are obtained by using lemma 7.3 and the inequality in (7.13).

We complete the proof of (i) by adding all the upper bounds for P_1 - P_4 and noting that $n_{1c} = O(c^{-1/(2+2\gamma)})$. The proof for part (ii) of lemma 7.4 is very similar to the proof of part (i). \square

Lemma 7.5. *If nonnegative i.i.d. observations X_1, \dots, X_n are such that $E(X_1^{4r})$ and $E(X_1^{-4r\alpha})$ exist for some $r \geq 1$ and $\alpha > 1$, then*

$$E\left(\max_{n_{1c} \leq n \leq n_{2c}} (G_n - G_F)^r\right) = O(n_{1c}^{-r/2}) \text{ as } c \downarrow 0.$$

Proof. Applying C_r inequality, we can write

$$\begin{aligned} (G_n - G_F)^r &= \left\{ \hat{\Delta}_n \left(\frac{1}{2\bar{X}_n} - \frac{1}{2\mu} \right) + \frac{1}{2\mu} (\hat{\Delta}_n - \Delta) \right\}^r \\ &\leq \frac{1}{2} \left\{ \hat{\Delta}_n^r \left(\frac{1}{\bar{X}_n} - \frac{1}{\mu} \right)^r + \frac{1}{\mu^r} (\hat{\Delta}_n - \Delta)^r \right\} \end{aligned}$$

By Cauchy-Schwarz inequality and lemma 9.2.4 of Ghosh et al. (1997), we have

$$\begin{aligned} &2E\left(\max_{n_{1c} \leq n \leq n_{2c}} (G_n - G_F)^r\right) \quad (7.15) \\ &\leq \left\{ E\left(\max_{n_{1c} \leq n \leq n_{2c}} \hat{\Delta}_n^{2r}\right) E\left(\max_{n_{1c} \leq n \leq n_{2c}} \left(\frac{1}{\bar{X}_n} - \frac{1}{\mu}\right)^{2r}\right) \right\}^{\frac{1}{2}} + \frac{1}{\mu^r} \left(\frac{r}{r-1}\right)^r E\left(\hat{\Delta}_{n_{1c}} - \Delta\right)^r \\ &\leq \left\{ E\left(\max_{n_{1c} \leq n \leq n_{2c}} \hat{\Delta}_n^{2r}\right) \right\}^{\frac{1}{2}} \left\{ \frac{1}{\mu^{2r}} E\left(\max_{n_{1c} \leq n \leq n_{2c}} (\bar{X}_n - \mu)^{4r}\right) E\left(\max_{n_{1c} \leq n \leq n_{2c}} \frac{1}{\bar{X}_n^{4r}}\right) \right\}^{\frac{1}{4}} + O(n_{1c}^{-r/2}). \end{aligned}$$

The last inequality is obtained by Cauchy-Schwarz inequality and lemma 2.2 of Sen and Ghosh (1981). Note that, by lemma 9.2.4 of Ghosh et al. (1997), lemma 2.2 of Sen and Ghosh (1981),

and existence of $E(X_1^{4r})$,

$$E \left(\max_{n_{1c} \leq n \leq n_{2c}} (\bar{X}_n - \mu)^{4r} \right) \leq \left(\frac{4r}{4r-1} \right)^{4r} E(\bar{X}_{n_{1c}} - \mu)^{4r} \leq O(n_{1c}^{-2r}), \quad (7.16)$$

$$E \left(\max_{n_{1c} \leq n \leq n_{2c}} \hat{\Delta}_n^{2r} \right) \leq \left(\frac{2r}{2r-1} \right)^{2r} E(\hat{\Delta}_{n_{1c}}^{2r}) < \infty, \text{ and} \quad (7.17)$$

$$E \left(\max_{n_{1c} \leq n \leq n_{2c}} \frac{1}{\bar{X}_n^{4r}} \right) \leq 1 + \int_1^\infty P \left(\max_{n_{1c} \leq n \leq n_{2c}} \frac{1}{\bar{X}_n^{4r}} \geq t \right) dt \leq 1 + \frac{E(\bar{X}_{n_{1c}}^{-4r\alpha})}{\alpha - 1} \quad (7.18)$$

is finite as $E(X_1^{-4r\alpha}) < \infty$. The last inequality is due to the maximal inequality for reverse submartingales (Lee, p. 112, 1990). Using (7.16)-(7.18) in the upper bound for (7.15), we complete the proof. \square

Lemma 7.6. *If $E(X_1^8)$ and $E(X_1^{-\alpha})$ exist for $\alpha > 8$, then $E \left[\sup_{n \geq m} V_n^2 \right] < \infty$ for $m \geq 4$.*

Proof. To prove lemma 7.6, it is enough to show that: (i) $E \left[\sup_{n \geq m} s_{wn}^2 \bar{X}_n^{-2} \right]$, (ii) $E \left[\sup_{n \geq m} \left| \frac{\hat{\Delta}_n}{\bar{X}_n^3} \hat{\tau}_n \right| \right]$, (iii) $E \left[\sup_{n \geq m} \frac{\hat{\Delta}_n^2}{\bar{X}_n^4} \right]$, and (iv) $E \left[\sup_{n \geq m} \frac{\hat{\Delta}_n^2}{\bar{X}_n^4} S_n^2 \right]$ are finite. Following Sen and Ghosh (p. 338, 1981), we have $E \left[\sup_{n \geq m} s_{wn}^4 \right] < \infty$ if $E[X_1^\alpha] < \infty$ for $\alpha > 4$ and $m \geq 4$. By (7.18), $E \left[\sup_{n \geq m} \bar{X}_n^{-4} \right] < \infty$ if $E[X_1^{-\alpha}] < \infty$ for $\alpha > 4$. Therefore,

$$E \left(\sup_{n \geq m} s_{wn}^2 \bar{X}_n^{-2} \right) \leq \left\{ E \left(\sup_{n \geq m} s_{wn}^4 \right) E \left(\sup_{n \geq m} \bar{X}_n^{-4} \right) \right\}^{1/2} < \infty.$$

For (ii), we note that $\hat{\Delta}_n$ and $\hat{\tau}_n$ are U-statistics. Using lemma 9.2.4 of Ghosh et al. (1997),

$$E \left(\sup_{n \geq m} |\hat{\Delta}_n^4| \right) \leq \left(\frac{4}{3} \right)^4 E(|\hat{\Delta}_m^4|) \text{ and } E \left(\sup_{n \geq m} |\hat{\tau}_n^4| \right) \leq \left(\frac{4}{3} \right)^4 E(|\hat{\tau}_m^4|).$$

Applying Cauchy-Schwarz inequality twice,

$$E \left(\sup_{n \geq m} \left| \frac{\hat{\Delta}_n}{\bar{X}_n^3} \hat{\tau}_n \right| \right) \leq \left\{ E \left(\sup_{n \geq m} |\hat{\Delta}_n^4| \right) E \left(\sup_{n \geq m} |\hat{\tau}_n^4| \right) \right\}^{\frac{1}{4}} \left\{ E \left(\sup_{n \geq m} |\bar{X}_n^{-6}| \right) \right\}^{\frac{1}{2}} < \infty,$$

if $E(X_1^8)$ and $E(X_1^{-\alpha})$ exist for $\alpha > 6$. Similarly, we can show that $E \left[\sup_{n \geq m} \frac{\hat{\Delta}_n^2}{\hat{X}_n^2} \right] < \infty$ if $E(X_1^4)$ and $E(X_1^{-\alpha})$ exist for $\alpha > 4$. Finally, $E \left[\sup_{n \geq m} \frac{\hat{\Delta}_n^2}{\hat{X}_n^4} S_n^2 \right] < \infty$ if $E(X_1^8)$ and $E(X_1^{-\alpha})$ exist for $\alpha > 8$. This completes the proof. \square

Lemma 7.7. *Let U_n be a U-statistics for estimating θ based on n observations. For any $\epsilon \in (0, 1)$,*

$$E \left(\max_{n_{2c} \leq n \leq n_c} (U_n - U_{n_c})^4 \right) = O \left(\frac{\epsilon}{n_c^2} \right) \text{ as } c \downarrow 0.$$

Proof. Since $\{U_n - U_{n_c}\}_{n=n_{2c}}^{n_c}$ is a reverse martingale, lemma 9.2.4 of Ghosh et al. (1997) yields

$$E \left(\max_{n_{2c} \leq n \leq n_c} (U_n - U_{n_c})^4 \right) \leq \left(\frac{4}{3} \right)^4 E (U_{n_{2c}} - U_{n_c})^4. \quad (7.19)$$

Let $V_n = U_n - \theta$. Using reverse martingale property of V_n , i.e., $E(V_{n_{2c}} | \mathcal{F}_{n_c}) = V_{n_c}$, we have

$$E(V_{n_{2c}} V_{n_c}^3) = E(V_{n_c}^4), \quad E(V_{n_{2c}}^3 V_{n_c}) \geq E(V_{n_c}^4), \quad \text{and} \quad (7.20)$$

$$E(V_{n_{2c}}^2 V_{n_c}^2) \leq \{E(V_{n_{2c}}^4) E(V_{n_c}^4)\}^{\frac{1}{2}} \leq E(V_{n_{2c}}^4). \quad (7.21)$$

Using (7.20)-(7.21) and asymptotic form of 4th central moment of U-statistics (Sen, p. 55, 1981),

$$\begin{aligned} E(U_{n_{2c}} - U_{n_c})^4 &= E(V_{n_{2c}}^4) + E(V_{n_c}^4) - 4E(V_{n_{2c}} V_{n_c}^3) - 4E(V_{n_{2c}}^3 V_{n_c}) + 6E(V_{n_{2c}}^2 V_{n_c}^2) \\ &\leq 7 \{E(V_{n_{2c}}^4) - E(V_{n_c}^4)\} = O \left(\frac{1}{n_{2c}^2} - \frac{1}{n_c^2} \right) + o \left(\frac{1}{n_c^2} \right) = O \left(\frac{\epsilon}{n_c^2} \right). \end{aligned} \quad (7.22)$$

(7.22) is obtained by noting that $n_{2c} = n_c(1 - \epsilon)$. Hence, the proof is complete. \square

Lemma 7.8. *If nonnegative i.i.d. observations X_1, \dots, X_n are such that $E(X_1^8)$ and $E(X_1^{-16\alpha})$ exist for $\alpha > 1$, then for $\epsilon \in (0, 1)$,*

$$E \left[\max_{n_{2c} \leq n \leq n_{3c}} (G_n - G_{n_c})^2 \right] = O \left(\frac{\sqrt{\epsilon}}{n_c} \right) \text{ as } c \downarrow 0.$$

Proof. $E \left[\max_{n_{2c} \leq n \leq n_{3c}} (G_n - G_{n_c})^2 \right] \leq E_1 + E_2$, where $E_2 = E \left[\max_{n_c \leq n \leq n_{3c}} (G_n - G_{n_c})^2 \right]$, and

$$\begin{aligned} E_1 &= E \left[\max_{n_{2c} \leq n \leq n_c} (G_n - G_{n_c})^2 \right] \\ &= E \left[\max_{n_{2c} \leq n \leq n_c} \left\{ \left(\frac{1}{\bar{X}_n} - \frac{1}{\bar{X}_{n_c}} \right) \frac{\hat{\Delta}_n}{2} + \frac{1}{2\bar{X}_{n_c}} (\hat{\Delta}_n - \hat{\Delta}_{n_c}) \right\}^2 \right] \leq \frac{E_{11} + E_{12}}{4}, \end{aligned} \quad (7.23)$$

where $E_{11} = E \left[\max_{n_{2c} \leq n \leq n_c} \left(\frac{1}{\bar{X}_n} - \frac{1}{\bar{X}_{n_c}} \right)^2 \hat{\Delta}_n^2 \right]$ and $E_{12} = E \left[\max_{n_{2c} \leq n \leq n_c} \frac{1}{\bar{X}_{n_c}^2} (\hat{\Delta}_n - \hat{\Delta}_{n_c})^2 \right]$. Applying Cauchy-Schwarz inequality thrice, we can write

$$E_{11} \leq \left\{ E \left[\max_{n_{2c} \leq n \leq n_c} (\bar{X}_n - \bar{X}_{n_c})^4 \right] \right\}^{\frac{1}{2}} \left\{ E \left[\max_{n_{2c} \leq n \leq n_c} \hat{\Delta}_n^8 \right] \right\}^{\frac{1}{4}} \left\{ E \left(\frac{1}{\bar{X}_{n_c}^{16}} \right) E \left(\max_{n_{2c} \leq n \leq n_c} \frac{1}{\bar{X}_n^{16}} \right) \right\}^{\frac{1}{8}}.$$

Using lemma 7.7, lemma 9.2.4 of Ghosh et al. (1997), (7.18), and the conditions of lemma (7.8), we conclude that $E_{11} = O(\sqrt{\epsilon}/n_c)$. Similarly, using Cauchy-Schwarz inequality and lemma 7.7, we have $E_{12} = O(\sqrt{\epsilon}/n_c)$. Therefore, $E_1 = O(\sqrt{\epsilon}/n_c)$. Following the same arguments as above, one can show that $E_2 = O(\sqrt{\epsilon}/n_c)$. Hence, lemma 7.8 is proved. \square

7.3. Proof of Theorem 3.1

The proof of the parts (i) and (ii) are similar to ?. (i) The definition of stopping rule N_c in (2.7) yields

$$\sqrt{\frac{A}{c}} V_{N_c} \leq N_c \leq m + \sqrt{\frac{A}{c}} (V_{N_c-1} + (N_c-1)^{-\gamma}). \quad (7.24)$$

Since $N_c \rightarrow \infty$ a.s. as $c \downarrow 0$ and $V_n \rightarrow \xi$ a.s. as $n \rightarrow \infty$, by theorem 2.1 of Gut (?), $V_{N_c} \rightarrow \xi$ a.s..

Hence, dividing all sides of (7.24) by n_c and letting $c \rightarrow 0$, we prove $N_c/n_c \rightarrow 1$ a.s. as $c \downarrow 0$.

(ii) Since $N_c \geq m$ a.s. and $n_c \geq 1$, dividing (7.24) by n_c yields

$$N_c/n_c \leq m + \frac{1}{\xi} \left(\sup_{c>0} V_{N_c-1} + (m-1)^{-\gamma} \right) \text{ almost surely,} \quad (7.25)$$

where $E \left(\sup_{c>0} V_{N_c-1} \right) < \infty$ by lemma 7.6. Since $N_c/n_c \rightarrow 1$ a.s. as $c \downarrow 0$, by the dominated convergence theorem, we conclude that $\lim_{c \downarrow 0} E(N_c/n_c) = 1$.

(iii) We need to show $\lim_{c \downarrow 0} R_{N_c}(G_F)/R_{n_c}^*(G_F) = \lim_{c \downarrow 0} (A/2cn_c)E(G_{N_c} - G_F)^2 + \frac{1}{2} \lim_{c \downarrow 0} E(N_c/n_c) = 1$. Thus, it is enough to show that $\lim_{c \downarrow 0} (A/cn_c)E(G_{N_c} - G_F)^2 = 1$, i.e., $\lim_{c \downarrow 0} n_c E(G_{N_c} - G_F)^2 = \xi^2$. Since we know that $n_c E(G_{n_c} - G_F)^2 = \xi^2$, it is sufficient to show that

$$\lim_{c \downarrow 0} n_c \{E((G_{N_c} - G_F)^2 - (G_{n_c} - G_F)^2)\} = 0. \quad (7.26)$$

Let $E_1 = E[(G_{N_c} - G_F)^2 I(N_c \leq n_{2c})]$. By (7.1), lemma 7.4, and lemma 7.5, we have

$$\begin{aligned} n_c E_1 &\leq E \left[\max_{n_{1c} \leq n \leq n_{2c}} (G_n - G_F)^2 I(N_c \leq n_{2c}) \right] \\ &\leq n_c \left\{ E \left[\max_{n_{1c} \leq n \leq n_{2c}} (G_n - G_F)^4 \right] P(N_c \leq n_{2c}) \right\}^{\frac{1}{2}} = O(c^h), \end{aligned} \quad (7.27)$$

where $h = (1 - 2\gamma)/(4 + 4\gamma) > 0$ using $\gamma \in (0, \frac{1}{2})$. Here, we assume that $E(X_1^{16})$ and $E(X_1^{-16\alpha})$ exist for $\alpha > 1$. Following the same arguments as in lemma 7.5, we can show that $E(G_{n_c} - G_F)^4 = O(n_c^{-2})$ provided $E(X_1^{16})$ and $E(X_1^{-16\alpha})$ exist for $\alpha > 1$. Let $E_2 = E[(G_{n_c} - G_F)^2 I(N_c \leq n_{2c})]$. By Cauchy-Schwarz inequality and lemma 7.4, we have

$$n_c E_2 \leq n_c \{E[(G_{n_c} - G_F)^4] P(N_c \leq n_{2c})\}^{\frac{1}{2}} = O\left(c^{\frac{1}{1+\gamma}}\right) \quad (7.28)$$

provided $E(X_1^{16})$ and $E(X_1^{-24})$ exist. Therefore, combining (7.27) and (7.28), we have

$$\lim_{c \downarrow 0} n_c E \left[\{(G_{N_c} - G_F)^2 - (G_{n_c} - G_F)^2\} I(N_c \leq n_{2c}) \right] = 0. \quad (7.29)$$

Using the same arguments as in lemma 7.5, one can show that $E \left[\max_{n \geq n_{3c}} (G_n - G_F)^4 \right] = O(n_{3c}^{-2})$ provided $E(X_1^{16})$ and $E(X_1^{-16\alpha})$ exist for $\alpha > 1$. Let $E_3 = E[(G_{N_c} - G_F)^2 I(N_c \geq n_{3c})]$.

Cauchy-Schwarz inequality and lemma 7.4 yields

$$n_c E_3 \leq n_c \left\{ E \left[\max_{n \geq n_{3c}} (G_n - G_F)^4 \right] P(N_c \geq n_{3c}) \right\}^{\frac{1}{2}} = O \left(c^{\frac{1}{2+2\gamma}} \right). \quad (7.30)$$

Following the same approach as in (7.28), $n_c E [(G_{N_c} - G_F)^2 I(N_c \geq n_{3c})] \leq O \left(c^{\frac{1}{2+2\gamma}} \right)$. Thus,

$$\lim_{c \downarrow 0} n_c E \left[\{ (G_{N_c} - G_F)^2 - (G_{n_c} - G_F)^2 \} I(N_c \geq n_{3c}) \right] = 0. \quad (7.31)$$

Hence, it remains to prove that

$$\lim_{c \downarrow 0} n_c E \left[\{ (G_{N_c} - G_F)^2 - (G_{n_c} - G_F)^2 \} I(n_{2c} \leq N_c \leq n_{3c}) \right] = 0. \quad (7.32)$$

Let $W = \{ (G_{N_c} - G_F)^2 - (G_{n_c} - G_F)^2 \} I(n_{2c} \leq N_c \leq n_{3c})$. Note that

$$\begin{aligned} W &= \{ (G_{N_c} - G_F) + (G_{n_c} - G_F) \} (G_{N_c} - G_{n_c}) I(n_{2c} \leq N_c \leq n_{3c}) \\ &\leq 2 \left\{ \max_{n_{2c} \leq n \leq n_{3c}} |G_n - G_F| \right\} \left\{ \max_{n_{2c} \leq n \leq n_{3c}} |G_n - G_{n_c}| \right\} I(n_{2c} \leq N_c \leq n_{3c}). \end{aligned}$$

Using Cauchy-Schwarz inequality, lemma 7.8, and following the lines of lemma 7.5,

$$\begin{aligned} n_c E[W] &\leq 2n_c \left\{ E \left(\max_{n_{2c} \leq n \leq n_{3c}} (G_n - G_F)^2 \right) E \left(\max_{n_{2c} \leq n \leq n_{3c}} (G_n - G_{n_c})^2 \right) \right\}^{\frac{1}{2}} \\ &\leq 2n_c \left\{ O(n_c^{-1}) O \left(\frac{\sqrt{\epsilon}}{n_c} \right) \right\}^{\frac{1}{2}} = O(\epsilon^{1/4}). \end{aligned} \quad (7.33)$$

Since (7.33) is true for any $\epsilon \in (0, 1)$, taking limit on both sides of (7.33) as $\epsilon \rightarrow 0$, (7.32) is proved. Hence, the proof of theorem 3.1 is complete.

7.4. Proof of lemma 2.1

By bivariate Taylor expansion of $f(\hat{\Delta}_n, 2\bar{X}_n) = \frac{\hat{\Delta}_n}{2\bar{X}_n}$ around $(\Delta, 2\mu)$,

$$\frac{\hat{\Delta}_n}{2\bar{X}_n} - \frac{\Delta}{2\mu} = \frac{\hat{\Delta}_n - \Delta}{2\mu} - \frac{\Delta}{2\mu^2}(\bar{X}_n - \mu) + R_{1n}, \quad (7.34)$$

where $R_{1n} = -2(\hat{\Delta}_n - \Delta)(\bar{X}_n - \mu)/b^2 + 4a(\bar{X}_n - \mu)/b^3$, $a = \Delta + p(\hat{\Delta}_n - \Delta)$, $b = 2\mu + p(2\bar{X}_n - 2\mu)$, and $p \in (0, 1)$. Let $E_{1n} = E(R_{1n}^2)$, $E_{2n} = \frac{1}{\mu}E\left(R_{1n}(\hat{\Delta}_n - \Delta)\right)$, and $E_{3n} = -\frac{\Delta}{\mu^2}E\left(R_{1n}(\bar{X}_n - \mu)\right)$. Squaring both sides of (7.34) and taking expectation,

$$E\left(\frac{\hat{\Delta}_n}{2\bar{X}_n} - \frac{\Delta}{2\mu}\right)^2 = \frac{1}{4\mu^2}V(\hat{\Delta}_n) + \frac{\Delta^2\sigma^2}{4n\mu^4} - \frac{\Delta}{2\mu^3}cov(\hat{\Delta}_n, \bar{X}_n) + \sum_{i=1}^3 E_{in}. \quad (7.35)$$

Using variance and covariance formulas for U-statistics (Lee, 1990), it is simple to show that $V(\hat{\Delta}_n) = \frac{4\sigma_1^2}{n} + O(n^{-2})$ and $cov(\hat{\Delta}_n, \bar{X}_n) = \frac{2}{n}(\tau - \mu\Delta)$. Therefore, it remains to show that $\sum_{i=1}^3 E_{in} = O(n^{-3/2})$. First, we work on E_{1n} . Note that $R_{1n}^2 = 4W_{1n} + 16W_{2n} - 16W_{3n}$, where $W_{1n} = \frac{(\hat{\Delta}_n - \Delta)^2(\bar{X}_n - \mu)^2}{b^4}$, $W_{2n} = \frac{a^2}{b^6}(\bar{X}_n - \mu)^4$, and $W_{3n} = \frac{a}{b^5}(\hat{\Delta}_n - \Delta)(\bar{X}_n - \mu)^3$. By Cauchy-Schwarz inequality and lemma 2.2 of Sen and Ghosh (1981),

$$E|W_{1n}I(\bar{X}_n > \mu)| \leq \frac{1}{16\mu^4}E\left(\left(\hat{\Delta}_n - \Delta\right)^2(\bar{X}_n - \mu)^2\right) = O(n^{-2}),$$

$$E|W_{1n}I(\bar{X}_n \leq \mu)| \leq \frac{1}{16}\left\{E\left(\left(\hat{\Delta}_n - \Delta\right)^4(\bar{X}_n - \mu)^4\right)E\left(\frac{1}{\bar{X}_n^8}\right)\right\}^{\frac{1}{2}} = O(n^{-2}),$$

provided $E(X_1^8)$ and $E(X_1^{-8})$ exist. Following the same approach, we have

$$\begin{aligned} E \left| W_{2n} I(\hat{\Delta}_n > \Delta) I(\bar{X}_n > \mu) \right| &\leq E \left((\bar{X}_n - \mu)^4 \frac{\hat{\Delta}_n^2}{(2\mu)^6} \right) = O(n^{-2}), \\ E \left| W_{2n} I(\hat{\Delta}_n > \Delta) I(\bar{X}_n \leq \mu) \right| &\leq E \left((\bar{X}_n - \mu)^4 \frac{\hat{\Delta}_n^2}{(2\bar{X}_n)^6} \right) = O(n^{-2}), \\ E \left| W_{2n} I(\hat{\Delta}_n \leq \Delta) I(\bar{X}_n > \mu) \right| &\leq \frac{\Delta^2}{(2\mu)^6} E (\bar{X}_n - \mu)^4 = O(n^{-2}), \\ E \left| W_{2n} I(\hat{\Delta}_n \leq \Delta) I(\bar{X}_n \leq \mu) \right| &\leq E \left((\bar{X}_n - \mu)^4 \frac{\Delta^2}{(2\bar{X}_n)^6} \right) = O(n^{-2}), \end{aligned}$$

provided $E(X_1^{12})$ and $E(X_1^{-18})$ exist. Similarly, we can show that $E(W_{3n}) = O(n^{-2})$ provided $E(X_1^{12})$ and $E(X_1^{-20})$ exist. Therefore, $E_{1n} = E(R_{1n}^2) = O(n^{-2})$. By Cauchy-Schwarz inequality and lemma 2.2 of Sen and Ghosh (1981), we obtain $E_{2n} = O(n^{-3/2})$ and $E_{3n} = O(n^{-3/2})$. Hence, lemma 2.1 is proved.

8. REFERENCES

- Aguirregabiria, V. and Mira, P. (2007), Sequential estimation of dynamic discrete games, *Econometrica*, 75, 153.
- Allison, P. D. (1978), Measures of inequality, *American Sociological Review*, 865880.
- Arcidiacono, P. and Jones, J. B. (2003), Finite mixture distributions, sequential likelihood and the em algorithm, *Econometrica*, 71, 933946.
- Beach, C. M. and Davidson, R. (1983), Distribution-free statistical inference with Lorenz curves and income shares, *The Review of Economic Studies*, 50, 723735.
- Chattopadhyay, B. and De, Shyamal, K. (2014a), Estimation Accuracy of an Inequality Index, *Recent Advances in Applied Mathematics, Modelling and Simulation*, Accepted.
- Chattopadhyay, B. and De, Shyamal, K. (2014b), Estimation of Gini Index within Pre-Specified Error, *Submitted*.
- Cochran, W. G. (1977), *Sampling techniques*, vol. 98, New York, Wiley and Sons.
- Dantzig, G. B. (1940), On the Non-Existence of Tests of Students Hypothesis Having Power Func-

- tions Independent of σ , The Annals of Mathematical Statistics, 11, 186192.
- Davidson, R. (2009), Reliable inference for the Gini index, Journal of econometrics, 150, 3040.
- Davidson, R. and Duclos, J.-Y. (2000), Statistical inference for stochastic dominance and for the measurement of poverty and inequality, Econometrica, 68, 14351464.
- Doob, J. L. (1953), Stochastic processes, New York Wiley.
- Gastwirth, J. L. (1972), The estimation of the Lorenz curve and Gini index, The Review of Economics and Statistics, 306316.
- Ghosh, B. K. and Sen, P. K. (1991), Handbook of sequential analysis, vol. 118, CRC Press.
- Ghosh, M. and Mukhopadhyay, N. (1979), Sequential point estimation of the mean when the distribution is unspecified, Communications in Statistics-Theory and Methods, 8, 637652.
- Ghosh, M. and Mukhopadhyay, N. (1981), Consistency and asymptotic efficiency of two stage and sequential estimation procedures, Sankhya: The Indian Journal of Statistics, Series A, 220227.
- Ghosh, M., Mukhopadhyay, N., and Sen, P. K. (1997), Sequential estimation, Wiley (New York).
- Greene, W. H. (1998), Gender economics courses in liberal arts colleges: Further results, The Journal of Economic Education, 29, 291300.
- Gut, A. (2009), Stopped random walks: Limit theorems and applications, Springer.
- Hoeffding, W. (1948), A class of statistics with asymptotically normal distribution, The Annals of Mathematical Statistics, 19, 293325.
- Hoeffding, W. (1961), The strong law of large numbers for U-statistics, Institute of Statistics mimeo series, 302.
- Hollander, M. and Wolfe, D. A. (1999), Nonparametric statistical methods, New York John Wiley and Sons.
- Kanninen, B. J. (1993), Design of sequential experiments for contingent valuation studies, Journal of Environmental Economics and Management, 25, S1S11.
- Lee, A. J. (1990), U-statistics: Theory and Practice, CRC Press.
- Lo'eve, M. (1963), Probability theory, Van Nostrand, Princeton, NJ.
- Loomes, G. and Sugden, R. (1982), Regret theory: An alternative theory of rational choice under

uncertainty, *The Economic Journal*, 92, 805824.

Mukhopadhyay, N. and De Silva, B. M. (2009), *Sequential methods and their applications*, CRC Press.

Robbins, H. (1959), Sequential estimation of the mean of a normal population. In *Probability and Statistics (Harold Cramer Volume)*, Almquist and Wiksell: Uppsala, pp 235245.

Sen, P. K. (1981), *Sequential nonparametrics: Invariance principles and statistical inference*, Wiley New York.

Sen, P. K. and Ghosh, M. (1981), Sequential point estimation of estimable parameters based on U-statistics, *Sankhya: The Indian Journal of Statistics, Series A*, 331344.

Sproule, R. (1969), A sequential fixed-width confidence interval for the mean of a U-statistic, Ph.D. thesis, Ph. D. dissertation, Univ. of North Carolina.

Xu, K. (2007), U-statistics and their asymptotic results for some inequality and poverty measures, *Econometric Reviews*, 26, 567577.